1. Using the formula $E \, Z = \int_0^\infty (1 - F_Z(t)) \, dt$ that holds for any RV $Z \geq 0$ (see lecture slide 71) and setting $X_a = X - a$ for brevity, one obtains, using the result of Problem 3(c) from PS-3, that

$$
E |X - a| = E X_a^+ + E X_a^- = \int_0^\infty (1 - F_{X_a}(t)) \, dt + \int_0^\infty (1 - F_{X_a}(t)) \, dt
$$

$$
= \int_0^\infty (1 - F_X(t + a)) \, dt + \int_0^\infty F_X(a - t) \, dt
$$

$$
= \int_a^\infty (1 - F_X(u)) \, du + \int_{-\infty}^{a} F_X(u) \, du \quad \text{[changing variables } u := \pm t + a]\}
$$

To minimise this expression in $a$, compute $\frac{d}{da}$ RHS and equate the result to zero: we obtain at $a$ that are continuity points of $F$:

$$
-(1 - F_X(a)) + F_X(a) = 0,
$$

which holds true when $F_X(a) = 0.5$, which means that $a$ is the median of $F$.

It may happen that the last equation doesn’t have solution: this occurs when there is a point $a_0$ s.t. $F(a_0-) < 0.5$ and $F(a_0) > 0.5$, so that $F$ has a jump at $a_0$. Note that, in such a case, one has $\frac{d}{da}$ RHS $< 0$ for all $a < a_0$ and $\frac{d}{da}$ RHS $> 0$ for all $a > a_0$, which means that the minimum of $E |X - a|$ is attained at $a_0$. But $a_0$ is clearly the median of $F$, right? Right.

2. For $Y := X^\alpha$, one has (using Thm 4.23 and $F_Y(y) = P(Y \leq y) = P(X \leq y^{1/\alpha})$)

$$
E X^\alpha = E Y = \int_0^\infty (1 - F_Y(y)) \, dy = \int_0^\infty (1 - F(y^{1/\alpha})) \, dy = \int_0^\infty (1 - F(x)) \, dx^\alpha,
$$

which results in the asserted formula for $E X^\alpha$, right?

3. The joint density of $(X_1, X_2, X_3)$ is, due to independence, given by the product

$$
f_{x_1}(x_1) f_{x_2}(x_2) f_{x_3}(x_3) = e^{-(x_1 + x_2 + x_3)} \quad (x_1 > 0, \quad j = 1, 2, 3). \quad \text{So one can just integrate that function over the region } \{x_1 \leq 2x_2 \leq 3x_3\}. \quad \text{The answer will be } \frac{18}{35} = 0.3(27).
$$

Alternatively, conditioning on the value of $X_2$ (this is just an integral version of the total probability formula; we haven’t discussed conditioning of that kind in detail yet, but it should make sense for you anyway):

$$
P(X_1 \leq 2X_2 \leq 3X_3) = \int_0^\infty P(X_1 \leq 2X_2 \leq 3X_3 | X_2 = x) \, P(X_2 = dx) \quad \text{[by independence]}
$$

$$
= \int_0^\infty P(X_1 \leq 2x \leq 3x) e^{-x} \, dx = \int_0^\infty P(X_1 \leq 2) P(X_3 \geq \frac{2}{3}x) e^{-x} \, dx
$$

$$
= \int_0^\infty (1 - e^{-2x}) e^{-2x/3} e^{-x} \, dx = \int_0^\infty e^{-5x/3} \, dx - \int_0^\infty e^{-11x/3} \, dx = \frac{3}{5} - \frac{3}{11} = \frac{18}{35},
$$

\[1\text{Formally speaking, at } a \text{ that is a jump points of } F, \text{ that will hold for the left and right derivatives at } a, \text{ which is still OK.} \]
4. Start by drawing a picture: \((X_1, X_2)\) lies on the (four straight) lines \(x_j = \pm 1, j = 1, 2\), w.p. \(\frac{1}{4}\) for each. That \(X_1X_2 = 0\) a.s. means that \((X_1, X_2)\) must always be on one of the coordinate axes. As each of the lines \(x_j = \pm 1, j = 1, 2\) intersects with coordinate axes at one point only, these intersection point \(((1,0), (0,1), (-1,0), (0,-1))\) must have probabilities \(\frac{1}{4}\) each.

Clearly, \(\mathbf{E}X_j = 0\) (either from the symmetry of the distribution or from a direct computation, as \(\mathbf{P}(X_j = \pm 1) = \frac{1}{2}\mathbf{P}(X_j = 0) = \frac{1}{4}\)), and \(\text{Var}(X_j) = \mathbf{E}X_j^2 - (\mathbf{E}X_j)^2 = \mathbf{E}X_j^2 = 0 \times \mathbf{P}(X_j = 0) + 1 \times \mathbf{P}(X_j = \pm 1) = \frac{1}{2}, j = 1, 2\), while \(\text{Cov}(X_1, X_2) = \mathbf{E}X_1X_2 - \mathbf{E}X_1\mathbf{E}X_2 = \mathbf{E}X_1X_2 = 0\), as \(X_1X_2 = 0\) a.s.

There is no independence as \(\mathbf{P}(X_1 = 0, X_2 = 0) = 0 \neq \frac{1}{4} = \mathbf{P}(X_1 = 0)\mathbf{P}(X_2 = 0)\).

5. (a) The m.q. error equals

\[\mathbf{E}(Y - aX - b)^2 = \mathbf{E}Y^2 + a^2\mathbf{E}X^2 + b^2 - 2a\mathbf{E}XY - 2b\mathbf{E}Y + 2ab\mathbf{E}X,\]

which is a quadratic form in \(a, b\). Differentiate to find the minimum point (the only stationary point must be the minimum, as the form is non-negative!):

\[
\begin{align*}
0 &= \frac{\partial}{\partial a} \cdots = 2a\mathbf{E}X^2 - 2\mathbf{E}XY + 2b\mathbf{E}X, \\
0 &= \frac{\partial}{\partial b} \cdots = 2b - 2\mathbf{E}Y + 2a\mathbf{E}X.
\end{align*}
\]

Substituting expressions for the moments in terms of the \(\mu\)’s, \(\sigma\)’s and \(\rho\), from the second equation one has \(b = \mu_Y - a\mu_X\), which turns the first one into

\[0 = a(\sigma_X^2 + \mu_X^2) - (\rho\sigma_X\sigma_Y + \mu_X\mu_Y) + (\mu_Y - a\mu_X)\mu_X = a\sigma_X^2 - \rho\sigma_X\sigma_Y.\]

So \(\hat{a} = \rho\sigma_Y/\sigma_X, \hat{b} = \mu_Y - \mu_X\rho\sigma_Y/\sigma_X\), yielding the best linear predictor of the form

\[\hat{Y} = \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) + \mu_Y.\]

(b) \(\mathbf{E}\hat{Y} = \rho\frac{\sigma_Y}{\sigma_X}\mathbf{E}(X - \mu_X) + \mu_Y = \mu_Y\), and \(\text{Var}(\hat{Y}) = (\rho\frac{\sigma_Y}{\sigma_X})^2\text{Var}(X - \mu_X) = \rho^2\sigma_Y^2\).

So \(\rho^2\) gives the portion of variability in \(Y\) that \(X\) can explain, using statisticians’ slang.

(c) As \(\mathbf{E}(Y - \hat{Y}) = 0\), one has

\[
\text{Cov}(X, Y - \hat{Y}) = \mathbf{E}(X - \mu_X)(Y - \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) - \mu_Y)
\]

\[
= \mathbf{E}(X - \mu_X)(Y - \mu_Y) - \rho\frac{\sigma_Y}{\sigma_X}\mathbf{E}(X - \mu_X)^2
\]

\[
= \rho\sigma_X\sigma_Y - \rho^2\sigma_X^2 = 0.
\]